Title: CASE-CONTROL VACCINE EFFECTIVENESS STUDIES: DATA COLLECTION, ANALYSIS AND REPORTING RESULTS

Authors: Jennifer R. Verani\textsuperscript{a}, Abdullah H. Baqui\textsuperscript{b}, Claire V. Broome\textsuperscript{c}, Thomas Cherian\textsuperscript{d}, Cheryl Cohen\textsuperscript{e}, Jennifer L. Farrar\textsuperscript{a}, Daniel R. Feikin\textsuperscript{a,f}, Michelle J. Groome\textsuperscript{g}, Rana A. Hajjeh\textsuperscript{a}, Hope L. Johnson\textsuperscript{h}, Shabir A. Madhi\textsuperscript{e,g}, Kim Mulholland\textsuperscript{i,j}, Katherine L. O'Brien\textsuperscript{i}, Umesh D. Parashar\textsuperscript{a}, Manish M. Patel\textsuperscript{a}, Laura C. Rodrigues\textsuperscript{i}, Mathuram Santosham\textsuperscript{i}, J. Anthony Scott\textsuperscript{i,k}, Peter G. Smith, Halvor Sommerfelt\textsuperscript{m,n}, Jacqueline E. Tate\textsuperscript{a}, J. Chris Victor\textsuperscript{o}, Cynthia G. Whitney\textsuperscript{a}, Anita K. Zaidi\textsuperscript{p}, Elizabeth R. Zell\textsuperscript{a}

Corresponding Author: Jennifer R. Verani; email: jverani@cdc.gov

Affiliations:

\textsuperscript{a} National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention; 1600 Clifton Rd. Atlanta, GA, USA
\textsuperscript{b} International Center for Maternal and Newborn Health, Johns Hopkins Bloomberg School of Public Health; 615 N Wolfe St, Baltimore, MD, USA
\textsuperscript{c} Rollins School of Public Health Emory University; 1518 Clifton Rd, Atlanta, GA USA
\textsuperscript{d} Department of Immunizations, Vaccines and Biologicals, World Health Organization; 20 Avenue Appia 1211 Geneva, Switzerland
\textsuperscript{e} Centre for Respiratory Diseases and Meningitis, National institute for Communicable Diseases; 1 Modderfontein Road, Sandringham, Johannesburg, South Africa
\textsuperscript{f} International Vaccine Access Center, Johns Hopkins Bloomberg School of Public Health; 615 N Wolfe St, Baltimore, MD, USA
\textsuperscript{g} Respiratory and Meningeal Pathogens Unit, University of Witwatersrand; Richard Ward, 1 Jan Smuts Ave, Braamfontein, Johannesburg, South Africa
\textsuperscript{h} Monitoring & Evaluation, Policy & Performance, GAVI Alliance; Chemin des Mines 2, 1202, Geneva, Switzerland
\textsuperscript{i} Murdoch Children's Research Institute; Royal Children's Hospital, 50 Flemington Rd, Parkville VIC 3052, Australia
\textsuperscript{j} Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Population Health , London School of Hygiene and Tropical Medicine; Keppel St, London WC1E 7HT, UK
\textsuperscript{k} KEMRI-Wellcome Trust Research Programme; P.O. Box 230-80108, Kilifi, Kenya
\textsuperscript{l} MRC Tropical Epidemiology Group; London School of Tropical Medicine and Hygiene; London, UK
\textsuperscript{m} Centre of Intervention Science in Maternal and Child Health and Centre for International Health, University of Bergen; P.O.Box 7800, Bergen, Norway
\textsuperscript{n} Department of International Public Health, Norwegian Institute of Public Health, PO Box 4404 Nydalen, Oslo, Norway
\textsuperscript{o} PATH; 2201 Westlake Avenue, Seattle, WA USA
\textsuperscript{p} Aga Khan University; Stadium Rd, Karachi, Pakistan
Abstract

The case-control methodology is frequently used to evaluate vaccine effectiveness post-licensure. The results of such studies provide important insight into the level of protection afforded by vaccines in a ‘real world’ context, and are commonly used to guide vaccine policy decisions. However, the potential for bias and confounding are important limitations to this method, and the results of a poorly conducted or incorrectly interpreted case-control study can mislead policies. In 2012, a group of experts met to review recent experience with case-control studies evaluating vaccine effectiveness; we summarize the recommendations of that group regarding best practices for data collection, analysis, and presentation of the results of case-control vaccine effectiveness studies. Vaccination status is the primary exposure of interest, but can be challenging to assess accurately and with minimal bias. Investigators should understand factors associated with vaccination as well as the availability of documented vaccination status in the study context; case-control studies may not be a valid method for evaluating vaccine effectiveness in settings where many children lack a documented immunization history. To avoid bias, it is essential to use the same methods and effort gathering vaccination data from cases and controls. Variables that may confound the association between illness and vaccination are also important to capture as completely as possible, and where relevant, adjust for in the analysis according to the analytic plan. In presenting results from case-control vaccine effectiveness studies, investigators should describe enrollment among eligible cases and controls as well as the proportion with no documented vaccine history. Emphasis should be placed on confidence intervals, rather than point estimates, of vaccine effectiveness. Case-control studies are a useful approach for evaluating vaccine effectiveness; however careful attention must be paid to the collection, analysis and presentation of the data in order to best inform evidence-based vaccine policies.

Key words: vaccines, case-control studies, evaluation studies
Introduction

New vaccines are licensed based on the results of randomized controlled trials demonstrating safety and efficacy. Yet even after licensure, there are often questions about how well a vaccine protects against disease in a “real world” context because of differences in epidemiologic contexts, host factors affecting immune response, vaccine implementation (e.g. varying dosing schedules), and the potential for waning immunity over time. The case-control method is commonly used to estimate effectiveness after a vaccine has been implemented in a public health system; recent examples include evaluations of vaccines against *Haemophilus Influenzae* type B (Hib), *Streptococcus pneumoniae*, influenza, rotavirus, and cholera. The results of case-control vaccine effectiveness studies can complement and extend the data generated by clinical trials.

However the potential for bias and confounding are important limitations to the case-control method. In 2012, a group of experts met to review recent experience with case-control studies evaluating the effectiveness of several vaccines; here we summarize the recommendations of that group regarding best practices for data collection, analysis and interpretation. (A separate paper provides an overview of the case-control method for evaluating vaccine effectiveness and reviews planning, design, and the identification and enrollment of cases and controls.) While case-control vaccine effectiveness studies have been carried out in countries of all income levels, this review focuses on their implementation in resource-poor settings.

Assessment of vaccination status

Vaccination status is the primary exposure of interest for case-control vaccine effectiveness studies, but it can be challenging to assess it accurately. Misclassification of vaccination status can affect the VE estimates in various ways. Non-differential misclassification of vaccination status (i.e. cases and controls have similar risks of misclassification) will bias the effectiveness estimate towards the
null. Differential misclassification (i.e. vaccine classification errors have different probabilities in cases and controls) can bias the effectiveness estimate towards or away from the null, or even result in a negative VE, giving the false impression that vaccinated are at greater risk of the target disease than unvaccinated. The same strategies to obtain vaccination history should be used for both cases and controls. Equal, intense effort must be made to obtain vaccination histories from all cases and controls, and those efforts should be clearly documented and reported.

Preferred sources of vaccination data are family-held vaccine records, clinic records, immunization registry data, or other written documentation of vaccines received and the dates on which they were administered. Doses not recorded on these documents are assumed to have not been received; although this assumption may be incorrect if recordkeeping is poor. Parent reporting of routine infant immunizations received, without written verification, may be unreliable. However, if parents report receipt of no vaccines of any type or receipt of only birth doses, such a history may be valid even in the absence of written confirmation since unvaccinated children rarely will have family-held records and generally parents are unlikely to state that the child is unvaccinated when in fact he or she did receive vaccines. Because excluding unvaccinated children will lead to bias, children with a parental report of having received no routine vaccines beyond birth doses should be included and considered to have received no doses of the vaccine of interest. All eligible cases and controls should be enrolled regardless of whether a documented vaccination history is available at the time of enrollment. Although those lacking a confirmed vaccination history (other than unvaccinated children) will be excluded from primary analyses because of missing data, the proportion of enrolled children for whom vaccination history could not be obtained should be described in the results, and sensitivity analyses used to assess the impact of missing data on the effectiveness estimates (see Analysis section).
Investigators should endeavor to understand factors associated with vaccination card availability and retention in the study setting, and whether those factors may also be linked to risk of disease or likelihood of vaccination. In preparation for the study, efforts can be made to improve availability of cards and/or the quality and completeness of data in the clinic records. If vaccine histories are unavailable for a sizeable proportion of children in the area (e.g. ≥5-10%), then efforts should be made to assess differences between children with and without documented histories. If important differences exist with regards with risk factors for disease, then a case-control study in that context is likely to yield biased effectiveness estimates. Case-control studies may not be a valid method for evaluating VE in settings where more than a small fraction of children lack a documented immunization history.

Abstracting vaccination data from family-held cards or clinic records is not always straightforward and can be a source of bias. Copies of the vaccination data source (e.g. digital photo, photocopies, or scanned images of the card or record) are extremely useful for controlling data quality. Copies can be used for double-abstraction (e.g. by two independent observers), which may improve the quality of data, particularly in settings where interpretation of information in the record may be challenging, for example, where parental-held records have no dedicated space for a new vaccine or for vaccines administered during campaigns. Copies potentially allow for blinding with regard to case or control status for the person abstracting the vaccination data. Vaccine lot numbers, if recorded, can aid in determining which vaccines were received. Dates of all relevant vaccine doses, including the vaccine of interest and other vaccines given on the same or similar schedules, should be carefully recorded.

**Other variables and unmeasured confounding factors**
In addition to vaccination status, data should be gathered on other variables that may confound the association between vaccination and the disease of interest. Known or hypothesized confounders should be identified before study initiation, accurately and thoroughly captured during data collection, and adjusted for in the analysis if they confound the association between vaccination and illness. As with all observational studies, some degree of unmeasured confounding often occurs in case-control studies and has the potential to substantially alter the measured VE. Unmeasured confounding may result from failure to collect data on a known confounder, insufficient or inadequate data collection for a known confounder, or lack of data on an unrecognized or unknown confounder.

A few strategies to quantify unmeasured confounders have been suggested. The first has been called a “bias-indicator” or “sham outcome” study. This is performed concurrently with a case-control study of vaccine effectiveness, where the effectiveness of the studied vaccine is measured against another disease which is not expected to be prevented by the vaccine. As the vaccine should confer no protection against this other disease, any measured vaccine effectiveness would be indicative of unmeasured confounding. A bias-indicator study of oral cholera vaccine in Mozambique evaluated the vaccine’s effectiveness against non-cholera diarrhea, and found an effectiveness of 35% (95% CI 18 to 65%); however after adjustment for known confounders the vaccine effectiveness was 0%. This suggests that while there was confounding of the effectiveness results, it was not due to unmeasured confounding. A limitation of the bias indicator study is the assumption that vaccine effects are specific to the vaccine target, whereas there is increasing evidence that some vaccines may have non-specific effects that could reduce the risk for non-targeted infections. Non-infectious illnesses (e.g. accidents or injuries) could be considered as outcomes for bias indicators studies. Another type of study to quantify unmeasured confounding has been dubbed a “sham exposure” or “sham case-control” study. Here vaccine effectiveness of another vaccine is measured against the disease of interest. In Kenya, investigators measured the effectiveness of diphtheria-tetanus-pertussis-Hib-
Hepatitis B vaccine against rotavirus disease among children prior to the expected introduction of the rotavirus vaccine in 2014 and found no protection. Because sham case-control studies are generally carried out before the introduction of a new vaccine, they require advance planning and resources. When feasible, they can be useful for planning case-control studies, for example by revealing the least biased control group or identifying measurable confounders in the population.

**Implementation and adherence to protocols**

The quality of data on enrollment, vaccination status, and potential confounders depends on writing and implementing clear protocols and Standard Operating Procedures (SOPs) for study conduct. Efforts to recruit cases and controls should be documented using standardized forms such as screening logs or registers; such documentation can be used to monitor the adherence to study procedures and identify lapses as quickly as possible.

Because of potential for selection bias in control enrollment for vaccine effectiveness case-control studies, it is particularly important to standardize, document clearly in logs, and regularly monitor at the field level, the process for enrolling controls. This should include the number of potential controls screened, number and timing of attempts made to enroll potentially eligible controls, the reasons for non-enrollment of potential controls, the frequency of refusals, and the number and characteristics of the controls who were not enrolled. Some methods for supervision of field staff enrolling controls may include GPS tracking of field staff (to monitor their locations and pace of recruitment and enrollment) and intermittent supervisor monitoring of the homes that were visited. Any departures from the protocol or SOPs must be reported to study lead investigators and documented.

**Analysis**
The statistical analysis of a case-control study for the evaluation of vaccine effectiveness should follow directly from the protocol and analysis plan, which should define the outcomes to be examined, as well as the exposures of interest (e.g. complete schedule, 2 or more doses). The “unadjusted” effectiveness from a case-control study is calculated as $(1 – \text{odds ratio for vaccination}) \times 100\%$.

For cases, vaccination status is defined based on the number of doses received before becoming ill and usually excludes doses received within the two weeks prior to allow for induction of immune response. For individually matched controls, a reference date should be defined in order to examine the control’s vaccination status before the corresponding case became ill\(^2\); the reference date is often based on the case’s date of illness onset, but may be based upon the date of hospitalization or sample collection. Doses received more than two weeks (if this is the period used for the case) before the reference date should be considered in the analysis. For frequency matched controls, the situation in which multiple controls are matched to multiple cases, there are different reference dates (or ages) associated with each of the cases and controls, and the analysis must take account of this. A method for doing this has been described by Keogh et al\(^5\).

The odds ratio is usually calculated from a logistic regression model, using unconditional logistic regression for unmatched or frequency matched studies, and conditional logistic regression for matched studies, with strata defined for each matched case-control set \(^5\). For simple conditional logistic regression, only discordant strata (e.g. vaccinated cases with at least one non-vaccinated control, or non-vaccinated case with at least one vaccinated control) contribute to the analysis\(^5\); thus in settings of very high or low vaccine coverage, the power of the analyses will be reduced.

While all efforts should be made in the study design phase to minimize confounding (e.g. by matching), it is usually necessary to also control for confounding in the analysis, where potential confounders are included as independent variables in a regression model. Because inclusion of multiple covariates can result in loss of statistical power, it is important to avoid including factors that are not
true confounders. There is no formal statistical test for evaluating whether to include a potential confounder in the final analysis. Some researchers approach the inclusion of confounders based on the past literature and include all potential confounders in a full model. Others prefer to evaluate potential confounders based on the data of the current study. A common approach to confounder evaluation is to include both vaccination status and single potential confounders, one at a time, as independent variables in the logistic regression model. If the OR associated with vaccination status changes by a predetermined, albeit arbitrary, percent (e.g. 10%) or more after adjusting for the potential confounder, then that variable is retained in the final multivariable model since it appears to impact the VE.

Another approach for determining which variables to include in a multivariable model is the use of directed acyclic graphs, which are causal diagrams used to identify a subset of covariates that address confounding while avoiding introduction of bias. Directed acyclic graphs have been used for case-control vaccine effectiveness studies of influenza. While different strategies for identifying important confounding variables are acceptable, the method used should be determined at the stage of developing the analytic plan.

Before deciding on a final model, some investigators prefer to examine whether the odds ratio (and thereby the VE) differ between strata of potential confounders (i.e. effect modification). This may be formally tested using appropriate interaction terms in the regression models. If such interaction is meaningful and statistically significant, stratum-specific VEs might be reported. For example, in a study of the 7-valent pneumococcal conjugate vaccine in the United States, the effectiveness against vaccine-type and non-vaccine type invasive pneumococcal disease was presented for healthy children and those with comorbidities, since this variable was found to have significant interaction with vaccination status.

Missing vaccination data present a problem in a vaccine effectiveness case-control study, since those with missing data likely differ from those with a documented vaccination history in ways that
could bias effectiveness estimates. One approach to handling missing vaccination histories is to conduct a sensitivity analysis. The simplest sensitivity analysis assumes those with a missing vaccination history are either all unvaccinated or all completely vaccinated, providing two estimates of effectiveness under two different assumptions. A study of the Hib vaccine conducted in the Dominican Republic used this approach and found very little impact on the results, suggesting that the findings of the primary analysis were not substantially biased by the missing vaccination history data. Sensitivity analysis could also be conducted to examine the impact of low (and potentially biased) enrollment of controls on effectiveness estimates by assuming a range of vaccine coverage for individuals who were eligible but not enrolled.

Methodological approaches to dealing with missing data have been advancing rapidly, and although there has been little work in vaccine effectiveness studies evaluating the usefulness of multiple imputation for missing vaccination histories for enrolled participants (or non-enrolled participants, as mentioned above), this approach warrants exploration. Nonetheless, all possible efforts should be made to obtain as complete information as possible on vaccination status of cases and controls; no sensitivity analysis or imputation can fully compensate for data completeness and validity.

**Reporting study results**

The STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) guidelines for reporting on case-control studies are an excellent reference for determining the key pieces of information to record for a vaccine effectiveness study. For case-control vaccine effectiveness studies, it is crucial to provide a clear and explicit description of the recruitment strategy for cases and controls, and to carefully document non-enrollment as well as enrollment. Readers should be given a clear understanding of how many potential cases and controls were screened to achieve the number of enrolled participants and the primary reasons for non-enrollment (e.g. not eligible, unable to contact, refused participation). The number of cases and controls with no documented vaccination history
should also be stated in the results. Relevant differences between included and not included cases and controls, as well as between those with and without reliable vaccination history, should be documented.

In interpreting the study findings, investigators should focus on the confidence intervals of effectiveness estimates. Although readers or policy makers may be naturally drawn to point estimates, confidence intervals add crucial information on the precision of these estimates. Reports of case-control vaccine effectiveness studies should also include a discussion of the limitations and potential sources of bias, taking into consideration the inherent limitations of the study design.

Conclusions

The case-control methodology is frequently used to evaluate the effectiveness of new vaccines, providing important data on the ‘real-world’ performance of vaccines that guide decisions about vaccine introduction and sustained use. However, the potential for bias and confounding is high, and can threaten the validity of the findings. Studies aimed at better understanding bias in case-control studies, such as a simulation model estimating potential biases in influenza vaccine effectiveness studies, can advance the field and provide more specific guidance regarding circumstances in which the case-control approach is likely to yield reliable results.

High quality vaccination data collected using the methods for cases and controls is crucial for vaccine effectiveness studies; in settings where documented vaccination histories are difficult to obtain, case-control vaccine effectiveness studies are unlikely to be useful. Variables that confound the association between vaccination and disease should be carefully measured and adjusted for in the analysis. In reporting the results of a case-control vaccine effectiveness study, it is important to include information that provides insight into the degree of possible bias in enrollment and data collection, such as the number of potential controls screened or the proportion of cases and controls with documented vaccine history. Vaccine effectiveness estimates should be presented with emphasis on the confidence
interval rather than the point estimate. In order for case-control studies to accurately guide vaccine policy decisions, data collection must be thorough and with careful attention to minimize bias, the analysis performed per the analytic plan with attention to potential confounding, and the results carefully interpreted and presented.

Acknowledgements

The authors would like to acknowledge Claudia DaSilva for organizing the meeting which formed the basis for this paper. We would also like to thank Dr. Jill Ferdinand and Tamara Pilishvili for their contributions to the scientific content of the meeting and subsequent discussions.

Author contributions

JRV: Conceptualization, Methodology, Writing: Original draft preparation Writing: Review and editing, Project administration, Supervision

AHB: Conceptualization, Writing: Review and editing

CVB: Conceptualization, Writing: Review and editing

TC: Conceptualization, Writing: Review and editing

CC: Conceptualization, Methodology, Writing: Review and editing

JLF: Writing: Review and editing, Project administration

DRF: Conceptualization, Writing: Original draft preparation, Writing: Review and editing

MG: Conceptualization, Writing: Review and editing

RAH: Conceptualization, Methodology, Writing: Review and editing, Funding acquisition

HLJ: Conceptualization, Writing: Review and editing

SAM: Conceptualization, Writing: Review and editing

KM: Conceptualization, Writing: Review and editing

KLO: Conceptualization, Methodology, Writing: Review and editing, Supervision
UDP: Conceptualization, Writing: Original draft preparation, Writing: Review and editing

MMP: Conceptualization, Writing: Original draft preparation, Writing: Review and editing

LCR: Conceptualization, Writing: Review and editing

MS: Conceptualization, Writing: Review and editing

JAS: Conceptualization, Methodology, Writing; Review and editing

PGS: Conceptualization, Methodology, Writing: Review and editing

HS: Conceptualization, Methodology, Writing: Review and editing

JET: Conceptualization, Writing: Review and editing

JCV: Conceptualization, Writing: Review and editing

CGW: Conceptualization, Methodology, Writing: Review and editing, Supervision

AKZ: Conceptualization, Writing: Review and editing

ERZ: Conceptualization, Writing: Review and editing

Funding: Funds from the GAVI Alliance covered the cost of an expert meeting held in November, 2012 to discuss the case-control method for evaluating vaccine effectiveness.

Disclaimers: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention. Thomas Cherian is a staff member of the World Health Organization. He alone is responsible for the views expressed in this publication, which may not necessarily represent the decisions or the policies of the World Health Organization.

Author disclosures of potential conflict of interest: CC reports having received grant funds from Sanofi Pasteur that were awarded to the National Institute for Communicable Diseases, South Africa
References


25. Castilla JÁ­s, Beristain X, MartÁ­nez-Artola Vc, NavascuÁ­ñ s a, GarcÁ­ñ a-Cenoz M, Alvarez N, Polo I, MazÁ­ñ aÁ­nh A, Gil Setas A, Barricarte A. Effectiveness of rotavirus vaccines in preventing cases and hospitalizations due to rotavirus gastroenteritis in Navarre, Spain. Vaccine 2012;30(3):539-543.


Mahoney RT, Maynard JE. The introduction of new vaccines into developing countries. Vaccine 1999;17(7-8):646-652.